AD-A134 050 PROCESSING PHENOMENA AND THE DISSOCIATION BETWEEN SUBJECTIVE AND OBJECTIV..(U) ILLINOIS UNIV AT URBANA ENGINEERING-PSYCHOLOGY RESEARCH LAB 1/1

UNCLASSIFIED M D VIDULICH ET AL. OCT 83 F/G 5/5 NL

END

FILMED

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

A134050

(12)

ENGINEERING-PSYCHOLOGY RESEARCH LABORATORY

University of Illinois at Urbana-Champaign

TECHNICAL REPORT EPL-83-2/ONR-83-2

OCTOBER 1983

Processing Phenomena and the Dissociation between Subjective and Objective Workload Measures

DTIC
ELECTE
OCT 26 1983
S D
B

Michael D. Vidulich
Christopher D. Wickens

Prepared for
Office of Naval Research
Engineering Psychology Program
Contract No. N-000-14-79-C-0658
Work Unit No. NR 196-158

DTIC FILE COPY

Approved for Public Release: Distribution Unlimited

83 10 011

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: EPL-83-2/ONR-83-2

2. GOVT ACCESSION NO.: A134050

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): Processing Phenomena and the Dissociation between Subjective and Objective Workload Measures

5. TYPE OF REPORT & PERIOD COVERED: Technical

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): Michael D. Vidulich, Christopher D. Wickens

8. CONTRACT OR GRANT NUMBER(s): N000-14-79-C-0658

9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Illinois, Department of Psychology, 603 E. Daniel St., Champaign, IL 61820

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 196-158

11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Engineering Psychology Program, 800 N. Quincy St., Arlington, VA 22217

12. REPORT DATE: October 1983

13. NUMBER OF PAGES: 40

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION STATEMENT A

Approved for public release

Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for public release. Distribution unlimited.

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

workload, subjective assessment, dissociation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Causes of dissociation between subjective workload assessments and objective performance were investigated. A Sternberg memory search task was utilized. Sternberg task configurations varied in the automaticity of performance, stimulus presentation rate, discernability of stimuli, and the value of good performance. Automaticity in Sternberg task performance was manipulated by using two independent sets of stimuli; one of which was consistently mapped (i.e., targets were always the same) while the other was-

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

con't

inconsistently mapped (i.e., targets changed over trials). Also, all Sternberg configurations were performed both as single tasks and as part of dual task combinations (with a manual control task). During testing subjects rated all trials on eight typical bipolar rating scales.

Analysis of the results detected three major differences (i.e., dissociations) between what the ratings of workload would predict and, the actual performance which occurred. Subjects' ratings: (1) did not reflect the dual-task advantage of the consistently mapped Sternberg, (2) predicted an advantage for the slower presentation rate in which performance was degraded, and (3) indicated a higher level of workload was associated with the performance gain in a bonus-available condition. All of these dissociations identified could potentially contaminate subjective assessments in the field. The results were interpreted as supporting cognitive-processing-based experimentation in subjective workload assessment aimed at identifying differences between the cognitive processing accounting for subjective assessments and those processes that produce performance.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/
Availability Codes
Dist | Avail and/or Special
A

Processing Phenomena and The Dissociation between Subjective and Objective Workload Measures

Workload measurement is one of the critical issues in engineering psychology. In many high performance man-machine systems the decision of whether or not to add, or how to configure, a potential subsystem is guided by the estimation of how much "workload" that subsystem would inflict on the operator. For example, the question of excessive workload was recently one reason for the recommended elimination of a nearly five billion dollar missile program ("Maverick Production Opposed by GAO," 1982).

Despite the importance and common usage of the workload concept there is no recognized definition of workload. This unsatisfying state of affairs may be at least partly due to the fact that workload is commonly considered to be multidimenstional (Johanssen, Moray, Pew, Rasmussen, & Wickens, 1979) and has generated a large variety of measurement methods. Each technique tends to make its own assumptions about the nature of workload, enjoy certain strengths, and suffer from certain weaknesses. The two most common workload assessments techniques are subjective assessments and objective (i.e., performance based) assessments.

Subjective assessment is the use of operator's opinion of how much workload he/she "feels" is being induced by performing a task. In practice the technique may consist of using only a few general non-standardized questions (e.g., "How difficult was that?") or may use more quantitative rating scales, such as the Cooper-Harper scale for aircraft handling qualities (Cooper & Harper, 1969). Subjective ratings are the most popular assessment methods. There are a number of reasons for this: First, the unintrusiveness of the technique is a distinct advantage. There are two major aspects to subjective assessments' lack of intrusiveness: (1) since they are usually collected retrospectively, rather than during action, they do not interfere with the operator's perception of the task environment, and (2) since they do not usually involve any recording concurrent with performance there is no need to interface recording equipment in what is often a crowded machine environment (e.g., single-seat aircraft cockpits). Second, the fact that subjective assessment requires no sophisticated recording equipment makes it a very economical procedure to use, in both time and money. Third, subjective assessments have a great deal of face validity. This is especially true to the operators themselves. The man on the spot is expected to best know the situation.

The second major category of workload assessment methods are the objective techniques. In this category are all techniques which are based on the observation of operators' performance. The most commonly used objective assessment technique is the spare mental capacity technique. The spare mental capacity technique usually incorporates a secondary task to be time-shared with the task being studied (primary task). The assumption is that performance on the secondary task will reflect the workload of the

primary task with lower workload being associated with better secondary task performance. After subjective ratings the secondary task technique is the most popular workload assessment method (Shingledecker, 1980). Objective measures are based on logical extrapolation from contemporary attentional theories and involve observable performance data which can be readily quantified. These factors make objective measures attractive to many potential users.

The overall emerging picture is that workload is an effect of increasing task demands that is estimated by changes in operator feelings or performance. In its simplest conception this idea would predict that increases in task demands should result in similar effects among the different categories of assessment techniques. There have been a number of cases in which two or more methods have been compared and this claim has been supported (e.g., Higgins, 1979; Bird, 1981).

But, recently there have also been a disconcerting number of cases in which the different methods have been found to indicate different levels of workload or, in other words, to "dissociate" from each other. Eggemeier, Crabtree, Zingg, Reid, and Shingledecker (1982) found that subjective ratings were more sensitive than objective error data to difficulty manipulations of a short term memory task, especially at low levels of task difficulty. Wickens and Yeh (1982) demonstrated three ways subjective and objective measures dissociate: (1) subjective measures are relatively more sensitive to increasing the number of concurrent tasks, (2) objective measures are relatively more sensitive to resource competition, and (3) increasing control order of a tracking task had a relatively greater effect on subjective ratings. Perhaps the most complete demonstration of dissociation are the findings of William Derrick's Ph.D. dissertation research (reported in Derrick, 1981; Wickens & Derrick, 1981). The research used representative measures from all three categories and found two major classes of dissociation: (1) subjective measures were found to be relatively more sensitive to the addition of tasks to be time-shared, whereas objective measures were relatively more sensitive to increasing single-task difficulty, and (2) subjective and objective measures were relatively more sensitive to resource competition between concurrent tasks, whereas a physiological measure, heart rate arrythmia, was relatively more sensitive to total resource demand.

The unfortunate, but common, reaction when such dissociations occur is to question one or more of the involved measures, especially the subjective measure. Perhaps as a postbehavioristic legacy there remains a tendency of psychologically-trained individuals to distrust or deride the value of what amounts to a form of introspective data. After all, the prime purpose of Human Factors work is to improve the performance of systems. If subjective ratings are not sensitive to factors that influence performance or are sensitive to factors that do not, then their utility to aid in reaching this goal may seem questionable.

However, it can be argued that as long as the different measures of workload are lawfully related to some aspect(s) of workload, then all can be productively employed by the human factors practitioner. This point of view leads to research explicitly concerned with investigating the dissociation between subjective ratings and observed performance. There are two reasons for this choice of concentration: First, subjective and objective measures are commonly used by applied personnel to make important desisions regarding man-machine system design. Clearly understanding the causes of dissociations between these measures should increase the validity of this work. Second, relevant theoretical concepts already exist concerning the relationship between the cognitive processes generating performance and those responsible for verbal reports, but are untested in the workload domain.

This state of affairs encourages a dissociation research strategy based on exploration of the theoretical cognitive processes which underlie subjective and objective workload measures, specifying in what ways they differ and determining where in practice these differences could result in a dissociation. Put another way, the goal is to link observed dissociations to theoretical cognitive processing phenomena. This can be referred to as a "processing-characteristic" approach.

So far, the dominant research approach in subjective workload assessment has been to attempt to catalog those aspects of task difficulty to which operator's subjective assessments are sensitive. This can be referred to as the "task-characteristic" approach. For example, Wewerinke and Smit (1974) used the Cooper-Harper scale and derivatives of it to test the relationship of subjective workload assessment to a manual control task of varying degrees of difficulty. Wewerinke and Smit (1974) concluded that the increases in subjective ratings were consistent with the objective estimate of the "control effort" predicted by the optimal control model. Higgins (1979) demonstrated a close relationship between force required to manipulate controls and the subjective difficulty of task performance in a flight simulator. Borg (1978) summarized a number of studies from his lab which suggested that subjective workload is related to the number of alternatives, insufficient data, uncertainty, inadequate feedback, time pressure, and perceived probability of failure.

All of these experiments and many more like them provide what could be important bits of information, IF interest centers in the same or very similar tasks. But finding a study, or combination of studies, to predict the reaction of subjective assessments in response to task demands for a novel task is difficult or impossible.

In contrast, the processing-characteristic approach suggests that changes in subjective assessments of difficulty should be linked to the properties of the theoretical cognitive processing associated with task performance. The expectation is that results based on these processing phenomena will transfer from studied to

novel task situations better than results based strictly upon objective task characteristics. Obviously, the technique is highly dependent upon the validity of the theoretical processing phenomena being examined. To be useful, processing phenomena being studied for its relevance to subjective workload assessment must be both well validated and generalizable.

For example, research by William Derrick (Derrick, 1981; Wickens & Derrick, 1981) can be considered a processing-characteristic study. Derrick explicitly selected tasks to manipulate the "resource competition" between combinations of tasks as predicted by the Wickens (1980) multiple resources model. Competition for resources is a hypothetical cognitive event which can generalize relatively easily to many real world situations that are considerably different from the ones studied in the experiment.

Of particular importance to processing-characteristic research is whether there is a difference between the cognitive processing responsible for objective performance and the processing responsible for verbal reports of the state of the processing system during performance. The existence and implications of such differences has been a topic of some interest to researchers completely outside the workload assessment area.

Verbal Reports as Data

For many years the value of verbal reports as psychological data has been debated. A classic confrontation occurred in the early part of this century with Watson's (1913) critique of analytic introspection as practiced by the Structuralist school (e.g., Titchner, 1912). However, even the champion of behaviorism found verbal reports, in the form of think-aloud protocols, an acceptable tool for some studies (Watson, 1920).

A modern resurgence of this debate started with a very discouraging analysis of verbal report utility by Nisbett and Wilson (1977). Nisbett and Wilson argued that "when people attempt to report on their cognitive processes. . . they do not do so on the basis of any true introspection. Instead, their reports are based on a priori, implicit causal theories, or judgments" (p. 231). Extended to the question of workload assessment, this would suggest that individuals asked to assess the workload generated by performing a task would do so on the basis of an a priori analysis of that task's difficulty rather than on the basis of any feelings of comfort or overload engendered concurrently with that task's performance.

However, a strong challenge to the Nisbett and Wilson position was advanced by Ericsson and Simon (1980). Ericsson and Simon (1980) adopt an information processing approach in which they analyze the processes responsible for generating verbal reports and how they relate to those processes which are responsible for performance. It is a rich information processing model they use, with one especially interesting aspect essential

for the present discussion; Ericsson and Simon (1980) suggest that:

> The important hypothesis for us is that due to the limited capacity of STM, only the most recently heeded information is accessible directly. However, a portion of the contents of STM are fixated in LTM before being lost from STM . . . We assume that any verbalization or verbal report of the cognitive process would have to be based on a subset of the information in these memories. (p. 223)

Translating this into more general terms, it can be asserted that the only processing events available to verbal report and therefore able to influence subjective workload assessments are those which are conscious or phenomenal and are either recent events that haven't been displaced from consciousness or are events which were successfully transferred to the more durable LTM storage.

Both the Nisbett and Wilson (1977) and the Ericsson and Simon (1980) viewpoints would agree that subjects' subjective assessments of workload may be based on what can be called a "logical" analysis. That is, assessments that are based on external analysis of the task characteristics, rather than on the subjective experience of performing the task. But, Nisbett and Wilson would argue that this is always the case since their model is predicated on the assumption that the important mental processess are unconscious, while Ericsson and Simon argue that at least some mental processes, particularly those in STM, are accessible to conscious verbal reports and that with proper methodology useful information can be gained. This would suggest that the accuracy of subjective workload assessments in predicting performance will vary with the nature of the cognitive processing involved in the performance being assessed.

Causes of Dissociation

Combining the Ericsson and Simon (1980) model of verbal reports and the problem of observed dissociations between subjective and objective workload measures leads to some potentially important insights. First, consider the potential effects of automaticity in task performance on subjective workload assessments. In complex real-world tasks, the overall performance is the result of a combination of numerous processes, some of which are automated and some of which are consciously controlled. What are the implications of mixing such phenomenally distinct processes to the dissociation of subjective and objective workload assessments? Certainly, if automatic processes typically have poor phenomenal representation, then it would be expected that their impact on subjective ratings of workload would be less accurate than conscious resource-limited processes in which effort, a very phenomenal component of performance, is a prime determinate of performance quality.

An experiment was performed to test this hypothesis. A modified Sternberg task was chosen as the task in which to manipulate automaticity. Manipulating the consistency with which items that can serve as targets can also serve as distractors in such a visual search paradigm has been demonstrated to greatly influence the development of automaticity in the performance of the task. Each subject had one set of stimuli which were consistently mapped; that is, stimuli which can serve as targets on some trials cannot occur as distractors on trials for which they are not targets. In this situation the subject's performance is expected to become, with practice, automated. For each subject another set of stimuli was used in a varied mapped procedure in which a letter which is a target on some trials is also likely to appear often as a distractor on trials for which it is not a target. In this situation automaticity will not develop. A number of other factors were manipulated along with consistency that were expected to aid in the evaluation of dissociations between workload measures.

The difficulty of the visual search task was manipulated in two ways in addition to the consistency manipulation: (1) perceptual loading, and (2) rate-changing. Perceptual loading of the visual search task was accomplished by covering the stimulus display area of the CRT with a cross-hatching of lines. Rate-changing the search task was accomplished by doubling the time between presentations and halving the number of test frames in a trial (therefore, this condition can also be referred to as the "slow" condition). Using such manipulations of task conditions was expected to provide a variety of effects in the performance and the subjective workload ratings of both the consistently mapped and the varied mapped search tasks.

Number of tasks to be performed was also manipulated. On half the test trials the visual search task time-shared with a manual-control tracking task. Despite the added complexity, there were three important reasons for including the tracking in this experiment. First, the tracking provided insurance against the possibility of ceiling effects in the varied-mapped, non-automated visual search. Second, the tracking should increase the challenge and make the trials more intrinsically motivating. Third, the tracking is similar to the manual control required in most vehicular control situations and thus increases the face validity of the experiment.

Finally, level of motivation was manipulated by offering payoffs for "good" performance on some of the test trials. "Good" performance was adjusted for each subject to an above average, but not impossible, level of performance in order to provide extra incentive. In dual-task trials the algorithm for determining payoffs was varied to influence the subject's priorities, either toward the search task or away from the search task (i.e., towards the tracking task).

This combination of tasks was expected to produce dissociations between subjective workload assessments and observed

performance. These dissociations can then be tied to the type of cognitive processing which produced them. Previous research (e.g., Wickens & Yeh, 1982) has demonstrated dissociations related to single-task/dual-task manipulations. In this experiment the effects of a variety of processing phenomena were examined in the single-task versus dual-task paradigm. The effect of automaticity is a particularly interesting and important question. If the assumption that automatic processes are essentially unconscious when in operation and therefore of no value in guiding verbal reports is correct then the subjects' workload assessments should be relatively inaccurate in the consistently mapped Sternberg.

The effect of bonus induced motivation and biasing on workload ratings and dissociations is an open question. Wickens and Yeh (1983) have suggested that increased motivation will improve performance and increase workload ratings. The single-task bonus/no-bonus manipulation in both the Sternberg and tracking tasks should provide tests of this. The effect of dual-task biasing on ratings and dissociations is a somewhat different question than in the single-task case and does not really relate to Wickens and Yeh's predictions. However, the effects of dual-task biasing on performance has been investigated often and an examination of its effects on subjective workload assessments is overdue.

The use of the perceptual-loading and the rate-changing manipulations was primarily a means of obtaining different levels of difficulty for the Sternberg task. Both manipulations have been demonstrated to affect Sternberg performance. Plus, the rate-change manipulation has been demonstrated to have profound effects on rating scales of the type used in this study (Hauser, Childress, & Hart, 1982).

Overall, the interaction of these variables should produce a useful data set for investigating some causes of dissociations between subjective workload assessments and objective performance.

Method

Subjects

Forty students of the University of Illinois were run in the experiment. Fifteen of the subjects were male, 25 were female.

Apparatus

Subjects were seated in a light and sound attenuated chamber. Both tasks were implemented on a PDP-11/40 computer. The computer was interfaced to a 10 cm x 8 cm CRT display via a Hewlett-Packard 1300 Graphics Display Interface. The display was about 90 cm in front of the subject and slightly below eye level. The subject's responses for the search task were accomplished through a three button control panel affixed to the right armrest of the chair. The buttons were pressed by the first three fingers of the subject's right hand. The buttons were 1 cm x 1 cm

square with the center button slightly offset in a forward direction. The subject's input for the manual control tracking task was via a MSI 521 joystick affixed to the left armrest of the chair. Subjects and the experimenter communicated through headphone and microphones.

Procedure

The typical subject started with two training sessions emphasizing the Sternberg task. Each Sternberg task trial, whether consistently or varied mapped, started by identifying the two target items followed by a set of probe displays. Each probe display consisted of two stimulus items presented in side by side boxes slightly below the center of the CRT screen. The visual display is portrayed in Figure 1. The letter search probe display portion consists of the boxes with the letters "Y" and "N." The subjects task consisted of indicating, as quickly as possible, the location of the target, either in the left or the right box. Twenty percent of the probe displays did not contain a target at all. On these probe displays the subject's task was to press the third button to indicate no target was present. Either target position or no target was indicated by pressing the appropriate button on the right keyboard. Each type of target/distractor mapping had a unique set of stimuli letters associated with it. For the consistent mapping condition the target letters were always "A" and "N" and the distractors were always "K," "S," "P," and "J." For the varied mapping condition the stimulus letters were "B," "C," "O," "E," "V," and "I." On any inconsistently mapped trial any two letters of the stimulus set could be the targets with the remaining four as the distractors.

On standard condition trials stimuli would appear in the two display boxes approximately every 1.5 seconds. Each trial consisted of 32 target-present trials and 8 no-target trials. The perceptual loading was accomplished by placing a cross-hatching of lines over the two search task display boxes. Rate-changing was accomplished by halving the number of probe displays to 20 (16 target-present, and 4 no-target) and doubling the ISI (i.e., increasing it to an average of 3.0 seconds). These manipulations were combined non-orthogonally with the consistent vs. varied manipulation, resulting in six single task search configurations.

The second task was a two-dimensional compensatory tracking task with velocity dynamics on the control stick and the display driven by a random forcing function with an upper cutoff frequency of .32 Hz. The display is illustrated in Figure 1. The crosshair was the target for the tracking task and the schematic aircraft was the cursor. The inner box indicates the extent of the space in which the cursor plane can "fly."

The first training session consisted of 18 practice trials of the consistently mapped search task with as many varied mapped search trials and single-task tracking trials as time permitted. Three of the trials were dual-task (i.e., both tasks were performed concurrently). At least two trials, but no more than

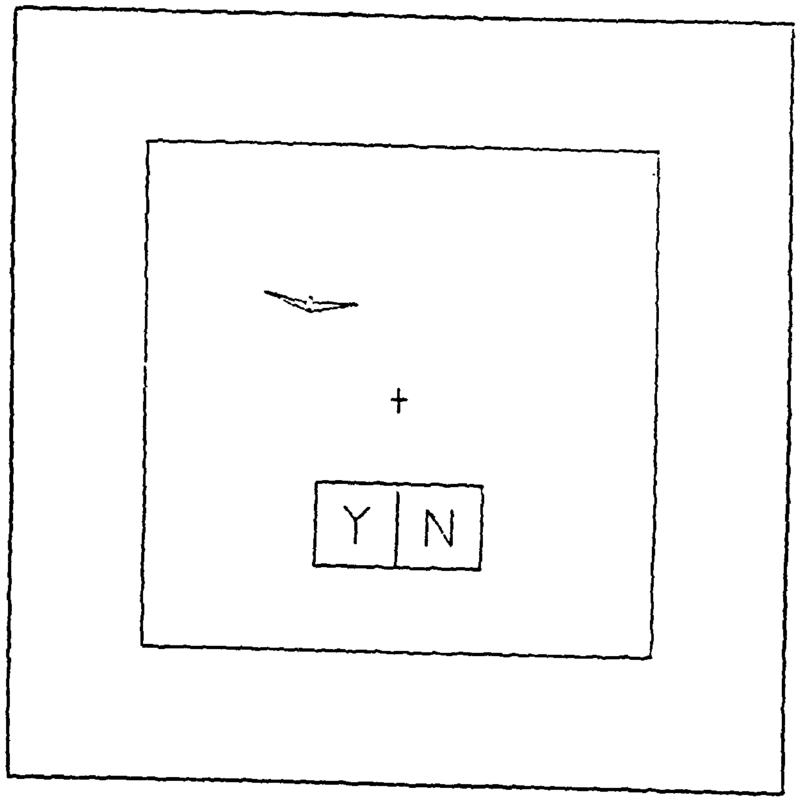

Figure 1. Dual-task display as seen on subject's CRT.

three, of each of the Sternberg difficulty manipulations were included. The second training session consisted of 10 consistently mapped search trials and as many varied mapped search trials and tracking trials as was necessary to stabilize performance. Single-task tracking root mean square (RMS) error had to be below .120 before performance was considered stabilized. Five trials were dual-task practice trials. One single-task and one dual-task trial was performed with each of the Sternberg difficulty manipulations. The last five or six trials of this session included collecting subjective assessments in order to familiarize subjects with the use of the scales. Each subject had between 800 and 900 opportunities to search for the consistent mapping targets before starting the test session.

The final session was the test session, in which each subject performed each single task configuration twice (once with payoffs available, once without), followed by 12 dual task trials. In these dual task trials payoffs were always available. But each dual-task configuration was run twice, once with a payoff strategy designed to favor the tracking and once with a payoff strategy designed to favor the search task. This means the total experimental block contained 26 trial types: 12 single-task Sternberg trials, 2 single-task tracking trials, and 12 dual-task trials.

During the test session each subject performed all 26 trial configurations once. All subjects performed the 14 single-task conditions prior to the 12 dual-task. Subjects in both groups received the 12 dual-task trials in random order.

The payoff criterion was adjusted for each subject as a function of their performance in the training trials. For each single task search task trial the subject was required to at least match their best percent accuracy score and improve upon their best RT score. In the single task tracking task they were required to beat their best overall RMS error score to earn the bonus. In the dual-task trials the pro-search task criterion was the same as the single task search task criterion with the addition that the subject at least be within .050 of their best training single task tracking overall RMS error score. In the pro-tracking conditions, the subject received the bonus for matching or beating their best single task overall RMS error score while coming at least within 10% accuracy and .10 second of their best single task search task score. Each bonus was worth 25 cents.

Following each trial on the test day the subject responded to a set of eight bipolar rating scales. The scales were a selection from Hauser, Childress, and Hart (1982) designed to test a variety of aspects of the subjective experience. The eight scales with their bipolar descriptors were: Overall Workload (very low, very high), Task Difficulty (very easy, very hard), Performance (very poor, very good), Mental/Sensory Effort (very low, very high), Response Load (very low, very high), Time Pressure (none, very rushed), Stress Level (relaxed, very tense),

and Incentive (very low, very important). Within this list there are two major categories of scale types: global and specific. The global scales (i.e., Overall Workload, Task Difficulty, and Performance) ask subjects to evaluate a number of attributes simultaneously. The specific scales (i.e., Mental/Sensory Effort, Response Load, Time Pressure, Stress Level, and Incentive) attempt to isolate certain aspects of the situation. Subjects were provided with a sheet of scale definitions to emphasize the differences between scales. These scales were selected because they have proven to be useful in previous research and are typical of the types of scales used by applied workers. Also, the relative usefulness of the global and specific scales is an important question.

The subjects' scale ratings were collected via the computer. Following each trial the computer would display the eight scales in sequence. Each display contained a scale title, a horizonal line with 14 slots marked on it, and the two endpoint descriptors. Above the line was a diamond shaped pointer. The subjects used the joy-stick to move the pointer to the appropriate slot on the horizonal line to indicate their ratings. The subjects then pulled the joy-stick's trigger and the computer would record the response and move on to the next scale, until all eight scales had been rated. The order the scales were presented in and the orientation of the endpoint descriptors were randomized.

In addition to the subjective ratings, each trial produced a variety of performance-based dependent measures. For the search task, accuracy and reaction time was averaged by probe types. For the manual control task, RMS error was averaged both over the whole trial and over the segments of time between a search display onset and the subject's response. These two classes of tracking measures will be referred to as overall or momentary RMS error.

Results

There were two major forms of data collected in this experiment: performance scores and subjective assessments. Ultimately, it is the relationship between these two classes of data which will be of primary interest in evaluating the experimental hypotheses. However, prior to examining this relationship it is necessary to first examine each independently. First, a review of the performance effects is in order. Testing the experimental hypotheses requires that the Sternberg consistency and difficulty manipulation were successful in producing the expected performance effects. Then, the ratings data will be studied. Two aspects of the data will be emphasized: changes in ratings over varying task conditions, and relationships between rating scales. Finally, the relationship between the performance effects and the subjective assessments will be examined for dissociations.

Performance Analysis

In the Sternberg task, whether consistently mapped or varied mapped, there are five distinct classes of responses possible. Assuming that there is a target letter present in the display the subject can: (1) correctly identify the target's position, (2) indicate the wrong position, or (3) fail to identify that there is a target at all and respond with the "no" button. These three response classes will be referred to as positive identifications, position errors, and misses. If, on the other hand, there is no target present in the probe display there are two types of responses the subject can make; a correct rejection, or a false alarm. Performance differences between the consistently mapped and the varied mapped trials across these classes of responses should provide a rich test of the presence and nature of automaticity of the consistently mapped trials.

Sternberg Latency Analysis. The correct response data was subjected to a pair of five-way ANOVAs. The five variables were: (1) number of task(s) (i.e., single-task or dual-task), (2) consistency (consistently mapped stimuli or varied mapped stimuli), (3) type of probe (target present or target absent), (4) pay, and (5) manipulation. The manipulation variable can refer to either the perceptual loading manipulation or the rate-changing manipulation. Since a non-orthogonal research design was used in the experiment, the separate analysis of each manipulation aids both the analysis and the interpretation of the data. The pay variable refers to the bonus manipulation: bonus availability in single-task trials and task bias in dual-task trials. This distinction must be considered when interpreting any effects involving the pay variable.

In the perceptual load analysis there were significant main effects for all five variables. Subjects were, on the average, 63 msec quicker in the single-task conditions, (F (1, 39) = 104.7, p < .0001). Responses in consistently mapped trials were also 60 msec faster than responses in varied mapped trials (F (1, 39) = 321.5, p < .0001). Perceptual-loading caused a 29 msec slowing in response time (F (1, 39) = 46.9, p < .0001). Pay exerted a significant influence on performance (F (1, 39) = 85.8, p < .0001). And finally, positive identifications averaged 483 msec versus 619 msec for the correct rejections (F (1, 39) = 756.5, p < .0001).

Two interactions were detected. The effect of the pay variable interacted with the number of tasks to be performed (F (1, 39) = 74.6, p < .0001). In single-task trials offering the bonus only improved performance by 3 msec (519 vs. 522 msec) while in the dual-task trials shifting the bias from the Sternberg task to the tracking task caused a 98 msec increase (i.e., from 533 msec to 631 msec). The consistency manipulation interacted with the type of probe (F (1, 39) = 64.8, p < .0001). Moving from the target present condition to the no-target condition increased the consistent mapping advantage. There were no other significant interactions in this analysis. The perceptual-loading

manipulation did not significantly interact with any other variable.

For the rate-changing manipulation the effects were much the same. Single-task responses were 58 msec faster than dual-task trials (F (1, 39) = 96.1, p < .0001). Responses in the consistently mapped trials averaged 510 msec while responses in the varied mapped trials took 65 msec longer (F (1, 39) = 196.9, p < .0001). The rate-changing manipulation slowed reaction times from the 537 msec average in the standard condition to 549 msec in the rate-changed condition (F (1, 39) = 18.2, p < .0001). Pay again exerted a significant effect (F (1, 39) = 55.0, p < .0001). Positive identifications took only 473 msec compared to the 612 msec required for a correct rejection (F (1, 39) = 902.0, p < .0001).

Time-sharing Sternberg performance with tracking produced a smaller decrement in the rate-changed condition than in the standard condition (49 vs. 65 msec; F (1, 39) = 7.2, p < .05). Once again, the presence of a bonus in the single-task condition produced only a modest improvement in response time (511 msec vs. 517 msec), whereas changing the bias from the tracking to the Sternberg task in dual-task conditions caused a much more pronounced improvement (617 msec vs. 525 msec; F (1, 39) = 95.2, p < .0001). The interaction between consistency and type of response was again significant (F (1, 39) = 39.1, p < .0001). Type of response had less of an effect on performance in the consistent condition. There were no other significant interactions in this analysis.

Sternberg Error Analysis. The proportions for each of the three error types (i.e., position error, miss, and false alarm) were calculated for each trial. These estimates of error probabilities were used in a pair of five-way ANOVAs comparable to those discussed in the last section. The only difference is that in this error analysis "type of error" is substituted for the variable "type of probe." Type of error has three levels corresponding to the three classes of possible errors.

The perceptual loading analysis identified two significant main effects: consistency (F (1, 39) = 88.0, p < .0001) and type of error (F (1, 39) = 128.1, p < .0001). Subjects were much less likely to commit an error in the consistently mapped trials (.018 vs. .042 for the varied mapped trials). Error likelihood was very close for the position errors (.015) and misses (.009), but much higher for the false alarms (.066).

Consistency and type of error also interacted with each other (F (2, 78) = 52.0, p < .0001). The consistent and inconsistent conditions are relatively close in error probability on target present trials, although the consistent condition had fewer errors. However, there is a large difference in performance on target absent trials. Subjects are much more likely to commit a false alarm in the varied mapped condition than in the consistently mapped condition.

In the analysis of the rate-changing manipulation the main effects directly paralleled those found in the previous analysis. Consistency and type of error were the only significant main effects (F (1, 39) = 43.1, p < .0001 and F (2, 78) = 61.6, p < .0001, respectively). The error probability on consistent trials was .019 while on varied trials it over doubled to .039. The error probabilities for the three types of errors were: .015 for position errors, .010 for misses, and .062 for false alarms.

There was an interaction between consistency and type of error (F (1, 39) = 19.9, p < .0001). This interaction is identical in form to the same interaction in the previous analysis. Most of the differences between the two consistency groups occurs in the false alarm response. Both consistency groups have higher probabilities of false alarms than position errors or misses, but the varied mapped trials have a much larger difference than the consistently mapped trials.

Tracking Performance

A set of two analyses was undertaken to contrast overall RMS error to momentary RMS error. Both analyses were four-way ANOVAs (Consistency x Perceptual-loading or Rate-changing x Bias x Type of RMS error Measure). In these analyses effects involving the type of measure variable are particularly important since these will isolate the effects of time-sharing relative to overall tracking performance.

Both analyses found significant main effects for the type variable (F (1, 39) = 104.0, p < .0001 in the perceptual-load analysis; F (1, 39) = 115.8, p < .0001 in the rate-change analysis). Both effects were results of higher error in the momentary RMS error than in the overall RMS error (.005 higher in the perceptual-load analysis; .009 higher in the rate-change analysis). This result is consistent with the expectation that the momentary RMS error isolates the periods of time when time-sharing is essential from those when tracking can be concentrated on. Apparently, at least some competition for resources occurs during the time-sharing period resulting in inflated RMS error scores for those periods.

Both analyses also displayed significant bias and type of measure interactions (F (1, 39) = 41.2, p < .0001 for the perceptual-load analysis; F (1, 39) = 11.8, p < .002 for the rate-change analysis). The increase in mean RMS error due to the momentary assessment technique is always less in the pro-tracking trials. This can be interpreted as evidence that the pro-tracking bias decreases the tendency to shift resources to the Sternberg task during Sternberg task stimulus presentation.

In the rate-change analysis there were two more significant interactions: Rate-changing x Type of Measure (F (1, 39) = 38.5, p < .0001), and Rate-changing x Bias x Type of Measures (F (1, 39) = 11.8, p < .002). At the heart of both of these interactions is a tendency for the rate-change manipulation to reduce overall RMS

error without affecting momentary RMS error. The Sternberg stimulus has equivalent disrupting effects in both the standard and the rate-changed conditions, but since there are fewer Sternberg stimuli in the rate-changed condition, the overall amount of disruption (and overall RMS error) is reduced. The three-way interaction involving bias displays this same basic tendency with the additional finding that the bias effect is identical in both the standard and the rate-changed conditions for the momentary RMS error measure while the overall RMS error measure shows a smaller bias effect in the rate-changed condition than in the standard condition. For the pro-tracking trials, the overall RMS error is roughly the same for both Sternberg task configurations (a difference of .001), but the rate-changed condition enjoys a relatively substantial advantage during pro-Sternberg trials (i.e., .011 less error). Again, this seems to be the result of the fact that in the slow presentation condition there are fewer Sternberg stimuli and, therefore, less disruption over the course of the trial. But, this mechanism is not important in the pro-tracking trials in which disruption effects are minimal anyway.

Summary of Performance Effects. The performance data indicate that the independent variables produced the expected differences in behavior. For example, there were decrements in the level of performance as subjects were required to perform two tasks simultaneously. The consistency variable substantially influenced performance in the expected directions. Subjects were faster and more accurate in the consistently mapped conditions. The consistent mapping also led to more stability in both reaction time and accuracy over the different classes of correct responses and error types. The bonus manipulation was relatively ineffective in the single-task conditions, but had a profound effect in the dual-task conditions. Both the perceptual-loading and rate-changing manipulations affected Sternberg performance, as well.

Ratings Analysis

The ratings data were analyzed in two ways. First, an ANOVA-based analysis procedure similar to that used in evaluating the performance effects was used. As in the previous analyses the effects of perceptual-loading and rate-changing were studied separately. To aid the interpretation of the significant main effects these analyses will be discussed in terms of their magnitude of effect. Second, multiple regression was used to investigate the relationship between individual global rating scales and combinations of specific item scales.

Task Effects on Ratings. Thus far this review of results has concentrated on F-tests of significance. Such analyses are concerned solely with detecting whether or not a treatment effect exists. In the present data there is an equally important question: How greatly does the magnitude of different treatment effects vary over dependent variables? In other words, an estimate of the importance of the independent variables in

determining the levels of the dependent variables is needed. As Myers (1979) has pointed out, "Neither the F ratio nor its level of significance provide this [i.e., an estimate of effect magnitude], since both these quantities are influenced by n and error variance" (p. 84). As one possible solution, Myers suggests the use of an estimate of the population absolute magnitude of effect. To generate such an estimate, Myers recommends subtracting the corresponding error mean squares from the mean squares associated with a significant effect and dividing by the number of subjects.

This approach was applied to the analysis of the ratings data. Two four-way Number of Tasks x Consistency x Pay x Perceptual-load or Rate-change ANOVAs were conducted on each of the rating scales. The magnitude of effect was calculated for all significant main effects. The results are displayed in Table 1. For the sake of comparison, the magnitude of the main effects of the reaction time analyses are included as well. For the rows associated with the number of tasks, consistency, and pay (single-task or dual-task) the data points represent mean magnitude of effect over both the perceptual-load and rate-change analyses. The data points in the perceptual-load and rate-change rows are based on only one analysis each, of course. Separate single-task and dual-task analyses were performed to provide the data for the pay rows; all other rows are based on analyses with both single-task and dual-task data included. A zero represents a non-significant result.

There are a number of general trends displayed on the table that are of interest. First, there is a very large difference in the average magnitude of effect between the effects in the reaction time measure and the subjective ratings effects. The average reaction time effect being much the larger. This indicates that the reaction time measure is much more sensitive than the rating scales. Whether this is due to a paucity of response categories or an impoverished phenomonal representation of task performance demands is an open question.

Comparing the magnitude of effects for the different independent variables on the rating scales, the most potent variable tends to be number of tasks. This contrasts somewhat with the RT data where the consistency variable has a greater effect. This replicates previous findings that have demonstrated the overwhelming effect of number of tasks in determining subjective assessments (e.g., Wickens & Derrick, 1981; Wickens & Yeh, 1982).

Two independent variables are notable for their lack of effect on subjective assessments: single-task pay and rate-changing. In the case of single-task pay this lack of effect on ratings is consistent with the lack of effect on Sternberg performance. Only the single-task Sternberg data are listed in the single-task pay row of Table 1. Some effects involving the single-task tracking will be reviewed later in this paper. However, the general inability of the rate-change manipulation to

Table 1

Magnitude of Significant Main Effects

I.V.s	OW	TD	PE	ME	RL	SL	TP	IN	RT
Number of Tasks	22.0	39.9	3.4	18.4	30.6	4.8	0	3.3	28212
Consistency	1.0	6.7	6.4	2.9	1.2	0.7	0	0	32551
Pay (ST)	0	0	0	0.5(S)	0	0	0.5(S)	5.6	0
Pay (DT)	0.6(P)	1.2(P)	2.7	0.5	0.4(P)	0	0.9(P)	0.5(S)	35431
Perceptual-load	2.2	6.5	0	2.6	0.9	0	0	0	6593
Rate-change	0	0	0	0	0	1.0	9.3	0.2	1135

Note. (S) = Significant only in rate-change analysis; (P) = Significant only in perceptual-load analysis; (ST) = Single-task; (DT) = Dual task; (OW) = Overall Workload; (TD) = Task Difficulty; (PE) = Performance; (ME) = Mental/Sensory Effort; (RL) = Response Load; (SL) = Stress Level; (TP) = Time Pressure; (IN) = Incentive

influence ratings is inconsistent with the relatively large magnitude of effect it had on Sternberg RT. Even more interesting is the fact that two of the scales that show a significant effect of rate-changing (i.e., Stress Level and Time Pressure) move in the opposite direction from what would be expected (i.e., they are rated easier in the condition with worse performance). These findings are indicative of a dissociation between measures. Further analyses involving this dissociation will be reviewed later.

Looking over the effects associated with the individual scales, differences in scale sensitivities can be detected. Overall, the most responsive scale was Task Difficulty which showed the largest magnitude of effect on three out of the four independent variables it responded to. The most disappointing scale is Time Pressure which responds to only three independent variables, two of which represent dissociations from the performance data.

There are three sets of interactions which bear mention. First, there were six occurrences of a significant Number of Tasks x Consistency interactions: four in the perceptual-load analysis (Task Difficulty, F (1, 39) = 38.6, $p < .0001$; Mental/Sensory Effort, F (1, 39) = 5.6, $p < .05$; Response Load, F (1, 39) = 8.9, $p < .005$; and Stress Level, F (1, 39) = 6.1, $p < .03$), and two in the rate-change analysis (Task Difficulty, F (1, 39) = 8.8, $p < .01$; and Mental/Sensory Effort, F (1, 39) = 5.2, $p < .05$). In every instance the means associated with these interactions showed a steeper rise in ratings moving from single-task to dual-task with the consistently mapped Sternberg task than in the varied mapped Sternberg task.

The effect of perceptual-loading on ratings interacted with number of tasks on four scales (Task Difficulty, F (1, 39) = 4.1, $p < .05$; Performance, F (1, 39) = 4.8, $p < .05$; Mental/Sensory Effort, F (1, 39) = 5.3, $p < .05$; and Response Load, F (1, 39) = 4.8, $p < .05$). Three of the Number of Tasks x Perceptual-loading interactions (Task Difficulty, Mental/Sensory Effort, and Response Load) reflected a larger increase in perceived workload as a result of the perceptual-loading manipulation in the single-task condition relative to the increase in the dual-task condition. The Performance scale Number of Tasks x Perceptual-load interaction was the result of subjects rating their dual-task performance higher in the standard condition but lower in the perceptually loaded condition.

The rate-change manipulation interacted with number of tasks (Overall Workload, F (1, 39) = 5.7, $p < .02$; and Time Pressure, F (1, 39) = 6.9, $p < .02$), consistency (Task Difficulty, F (1, 39) = 4.3, $p < .05$; and Mental/Sensory Effort, F (1, 39) = 6.4, $p < .02$), pay (Response Load, F (1, 39) = 4.4, $p < .05$), and number of tasks and consistency (Response Load, F (1, 39) = 7.8, $p < .01$; and Stress Level, F (1, 39) = 4.9, $p < .05$). Both number of tasks and rate-change interactions resulted from a larger increase in ratings going from single-task to dual-task in the rate-changed

(i.e., slower) task than in the standard task. For the Task Difficulty scale changing from consistent to varied mapping produced a larger increase in ratings in the standard condition than in the rate-changed condition; while for the Mental/Sensory effort scale the consistency and rate-change interaction reversed and the larger increase was associated with the rate-changed condition. The availability of a bonus lowered Response Load ratings in the standard condition, but raised them in the rate-changed condition. In both the Response Load and the Stress Level ratings, the three-way Number of Tasks x Consistency x Rate-change interactions reflected the fact that differences in ratings between the consistent and varied mapped conditions were uniformly small in the single- and dual-task rate-changed conditions, and the dual-task standard condition (i.e., between -0.2 units and 0.4 units), but were over 1 unit higher in the varied mapped single-task standard condition than in the corresponding consistent condition.

Predicting Global Ratings. Table 2 shows the results of a set of multiple regressions predicting individual global ratings from combinations of specific rating scales. The same data set containing 24 observations per subject used in calculating rating intercorrelations was used in this analysis. A total of nine multiple regressions were performed, three for each of the three global scales: Within each set of three multiple regressions the first equation is based on the overall data (i.e., including both consistent and varied mapped Sternberg trials). The second equation of each set was calculated using the data from only the trials employing consistently mapped Sternbergs. The third equation is based on the data from varied mapped Sternberg trials.

For both the Overall Workload scale and the Task Difficulty scale, regardless of the data set used, only three of the five specific scales were found to significantly contribute to explaining the variance in the global scales. In all six cases the same three scales were identified: Mental/Sensory Effort, Response Load, and Stress Level. The six equations listed in Table 2 using combinations of these three scales could account for between 49 and 61 percent of the global scale variance.

All three of the equations involving the Performance Scale value as the criterion variable found only two specific scales which could contribute significantly to explaining global scale variance. The two specific scales were Stress Level and Incentive. But, even at best, these two scales explain only 4 to 5 percent of the performance scale variability.

Overall, these results seem to imply a close relationship between the specific scales and the general experience of workload or task difficulty. But, the specific scales apparently do not tap the factors that influence the subjects' evaluations of their performance.

Table 2

Results of Multiple Regression Analysis Predicting Global Scale Values from Specific Scales

Criterion Scale and Data Set	Equation	Multiple R
<u>Overall Workload</u>		
Overall	Y = 1.2 + 0.37(ME) + 0.29(RL) + 0.19(SL)	.75
Consistent Set	Y = 1.2 + 0.37(ME) + 0.32(RL) + 0.16(SL)	.78
Varied Set	Y = 1.4 + 0.37(ME) + 0.26(RL) + 0.22(SL)	.72
<u>Task Difficulty</u>		
Overall	Y = -0.5 + 0.33(ME) + 0.37(RL) + 0.30(SL)	.74
Consistent Set	Y = -0.8 + -.46(RL) + 0.28(SL) + 0.28(ME)	.76
Varied Set	Y = 0.3 + 0.34(ME) + 0.32(SL) + 0.29(RL)	.71
<u>Performance</u>		
Overall	Y = 8.3 - 0.24(SL) + 0.16(INC)	.23
Consistent Set	Y = 8.3 - 0.22(SL) + 0.19(INC)	.23
Varied Set	Y = 8.1 - 0.23(SL) + 0.12(INC)	.20

Note. (ME) = Mental/Sensory Effort rating, (RL) = Response Load rating, (SL) = Stress Level rating, (INC) = Incentive rating.

Dissociation Analysis

In this final section of the results the relationship between the subjects' subjective workload assessments and their objective performance will be examined. The z score analysis procedure developed by Wickens and Yeh (1982) will be employed. First, both representative performance scores and global rating scores will be transformed subject-by-subject to z-scores. Then the z-scores for performance and the z-scores for ratings will be entered into an ANOVA as two levels of an independent variable. This variable will be referred to as "type of measure." Interactions between type of measure and any other independent variable(s) could indicate a dissociation between the measures in their sensitivity to the other independent variable(s). This procedure was employed by Wickens and Yeh (1982) to demonstrate a number of dissociations.

Z-score Analysis. For each subject, z-scores were calculated for the three global scales (i.e., Task Difficulty, Overall Workload, and Performance) and two performance measures (i.e, correct reaction time to target-present trials and momentary RMS error). In generating the z-scores for the reaction time measure analyses, data from both single-task and dual-task trials were included (except for the single-task tracking trials). The z-scores for the momentary RMS error analyses utilized only dual-task performance and rating data. In either case, each subject's mean was subtracted from individual observation and the difference divided by that subject's standard deviation of scores. The logic of this analysis technique is that when the performance and ratings measures are both converted to z-scores, the means and standard deviations are made to be equal (i.e., 0 and 1 respectively). However, the z-score transformation technique does not change the ordering of the different conditions within each measure type. Therefore, an ANOVA performed on this data with one set of ratings z-scores and one set of performance z-scores as two levels of one independent variable, referred to as type of measure, along with the variables associated with the experimental manipulations should detect dissociations. Dissociations will result in interactions involving the type of measure variable.

The reaction time z-score data were subjected to a set of five-way (Number of Tasks x Consistency x Perceptual-loading or Rate-changing x Pay x Type of Measure) ANOVAs. For both the perceptual-loading and the rate-changing manipulations three ANOVAs were performed; one comparing each global scale to reaction time performance. Three sets of interesting interactions involving the type of measure variable were detected.

The Number of Tasks x Consistency x Type of Measure interactions were significant in the Task Difficulty scale of both the perceptual-load (F (1, 39) = 20.8, $p < .0001$) and the rate-change (F (1, 39) = 11.2, $p < .002$) analyses. This interaction was also significant in the Overall Workload scale of

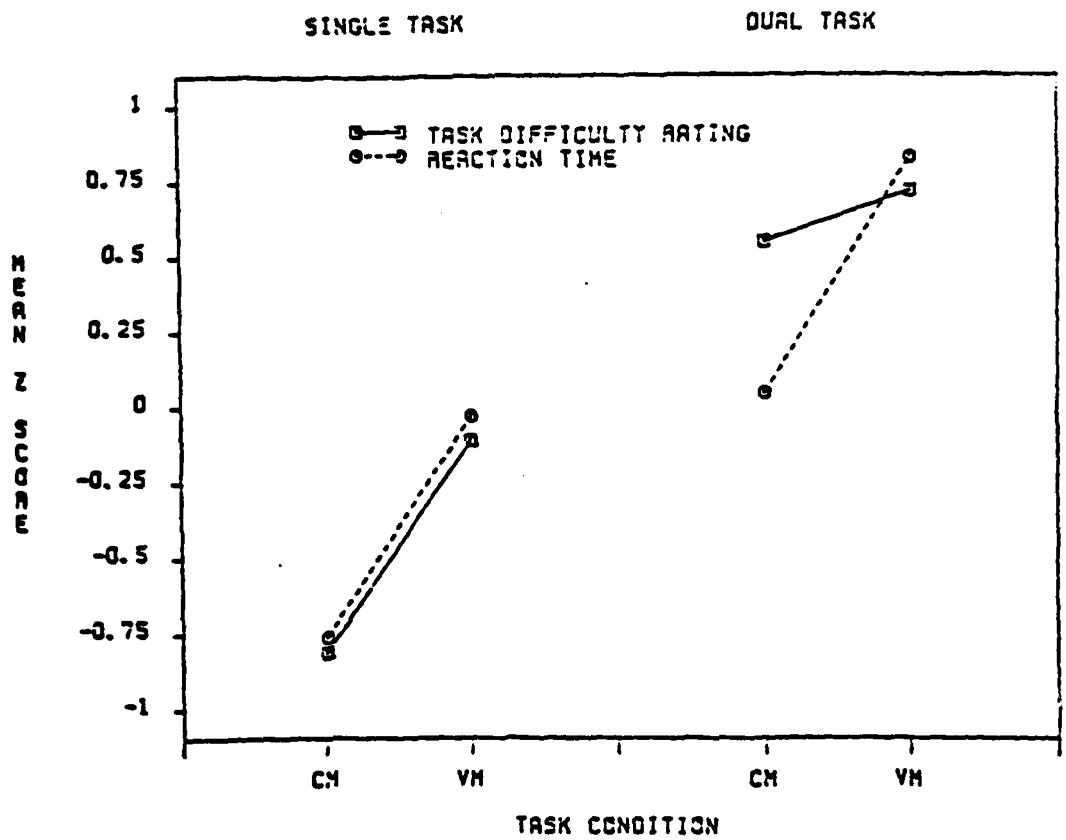

Figure 2. Perceptual-load z-score analysis Number of Tasks x Consistency x Type of Measure interaction. The measures involved are Task Difficulty rating and reaction time.

the perceptual-load analysis (F (1, 39) = 5.0, $p < .05$). An example of this interaction is displayed in Figure 2. In all three cases the interaction seems to be primarily a result of subjects' ratings of the dual-task with the consistently mapped Sternberg indicating a much higher level of workload or difficulty than would be expected from the reaction time to the Sternberg. Basically, the presence of the tracking tends to wipe out distinctions between the Sternberg configurations. Clearly, this is a potentially important finding. Certainly, it is important in applied settings in which one task in a multi-task environment is of primary interest. However, the theoretical interpretation is somewhat less straight-forward. The performance scores in these interactions are based solely on Sternberg reaction time data. Obviously, in the dual-task trials the tracking task performance is relevant to the subject's experience. For the most part, tracking performance was unaffected by the consistency of the Sternberg task. Consequently, if the tracking data were plotted on Figure 2, they would appear as a relatively horizonal line near the center of the right panel. Since the tracking task is a continuous task, as opposed to the Sternberg being discrete, the subjects' subjective experience could be more influenced by the tracking task. The ratings could possibly represent an accurate averaging of the two tasks' difficulty with the tracking task weighted more. Therefore, in a theoretical sense, there may be no dissociation occurring. Nevertheless, in applied settings in which one task's configuration is being manipulated in a multi-task environment this mechanism could produce misleading dissociations.

The second set of interactions involve the rate-change manipulation. The means associated with these interactions are displayed in Table 3. Rate-changing interacted with type of measure on both the Task Difficulty (F (1, 39) = 7.0, $p < .05$) and the Overall Workload (F (1, 39) = 9.2, $p < .01$) scales. The slower rate-changed Sternberg reduced ratings of Task Difficulty and Overall Workload but increased reaction time.

The Task Difficulty and Overall Workload scales also showed interactions between number of tasks, rate-changing, and type of measure (F (1, 39) = 5.5, $p < .05$; and F (1, 39) = 14.6, $p < .0005$;respectively). The Task Difficulty interaction is displayed in Figure 3. These interactions indicate that the locus of the rate-change and type of measure interaction is in the single-task condition. In the single-task condition the rate-changed Sternberg receives lower Overall Workload and Task Difficulty ratings, but also shows a slowing in reaction time. In the dual-task case rate-changing has little effect on either ratings or performance. These opposite trends are a very strong example of dissociation.

The last set of interactions involve the pay variable. All three scales in both manipulation analyses showed significant Number of Tasks x Pay x Type of Measure three-way interactions. (In the perceptual-load analysis: Task Difficulty, F (1, 39) = 37.4, $p < .0001$; Overall Workload, F (1, 39) = 23.9, $p < .0001$;

Table 3

Mean $\underline{Z}$-Score Data for Rate-Change x Type of Measure Interactions

Measure Type	Standard Condition Mean	Rate-Changed Mean
Task Difficulty Rating	-0.13	-0.17
Overall Workload Rating	-0.07	-0.16
Reaction Time	-0.23	-0.05

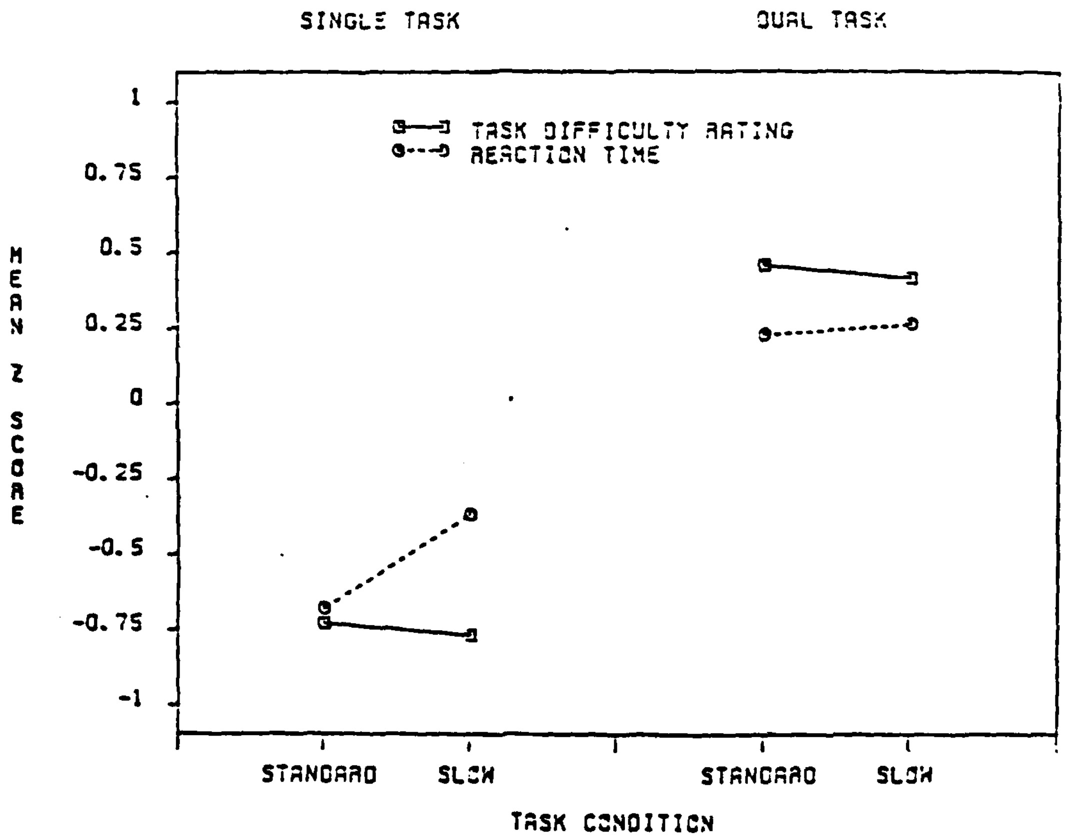

Figure 3. Rate-change x Number of Tasks x Type of Measure interaction in the z-score dissociation analysis. The measures involved are Task Difficulty rating and reaction time.

Table 4

Number of Tasks x Pay x Type of Measure Interaction Data

Analysis and Measure Type	Single Task No Bonus	Single Task Bonus	Dual Task Pro-TR	Dual Task Pro-TR
Perceptual-Load				
Task Difficulty	-0.51	-0.41	0.78	0.49
Overall Workload	-0.51	-0.33	0.71	0.44
Performance	0.14	0.19	-0.06	-0.38
Reaction Time	-0.38	-0.40	1.14	-0.26
Rate-Change				
Task Difficulty	-0.79	-0.70	0.52	0.37
Overall Workload	-0.75	-0.60	0.53	0.35
Performance	0.10	0.20	0.18	-0.38
Reaction Time	-0.48	-0.57	0.90	-0.41

Performance, $\underline{F}$ (1, 39) = 26.2, $\underline{p} < .0001$. In the rate-change analysis: Task Difficulty, $\underline{F}$ (1, 39) = 41.5, $\underline{p} < .0001$; Overall Workload, $\underline{F}$ (1, 39) = 23.0, $\underline{p} < .0001$; Performance, $\underline{F}$ (1, 39) = 7.9, $\underline{p} < .05$.) Table 4 displays the pertinent data for all 6 interactions. In the single-task conditions all of the scales show a slight increase in the bonus-available condition, while the reaction time shows a slight decrease. In the dual-task conditions both ratings and reaction time were reduced in the pro-Sternberg condition but the reaction time was reduced much more sharply.

The dual-task ratings data and RMS error data were run through a parallel set of six analyses. The only difference is that these data, having no number of tasks variable, were analyzed using a four-way ANOVA (i.e., Consistency x Pay x Type of Measure x Perceptual-loading or Rate-changing). There were six statistically significant results of interest.

Consistency and type of measure interacted in every analysis except the perceptually-loaded Overall Workload scale analysis. The means for the five interactions are reported in Table 5. In the two interactions involving Task Difficulty (perceptual-load analysis, $\underline{F}$ (1, 39) = 5.7, $\underline{p} < .03$; rate-change analysis, $\underline{F}$ (1, 39) = 15.5, $\underline{p} < .003$) and the one interaction involving Overall Workload in the rate-change analysis ($\underline{F}$ (1, 39) = 5.4, $\underline{p} < .03$) the interactions result from an increase in the ratings being combined with no real change in the RMS error score. In the two interactions involving the Performance scale (perceptual-load analysis, $\underline{F}$ (1, 39) = 12.5, $\underline{p} < .001$; rate-change analysis, $\underline{F}$ (1, 39) = 7.8, $\underline{p} < .01$) there is a decrease in the ratings combined with the negligable changes in the RMS error score. These results are consistent with the expectation of increased dual-task interference with the varied mapped Sternberg as opposed to the consistently mapped Sternberg. However, there are no performance effects which correspond to these ratings effects.

A three-way interaction between rate-change, pay, and type of measure from the rate-change analysis ($\underline{F}$ (1, 39) = 4.4, $\underline{p} < .01$) is displayed in Figure 4. Tracking performance shows approximately the same drop in performance as a result of a pro-Sternberg bias in both the standard and the rate-changed (slow) conditions. However, in the standard condition the ratings of task difficulty are reduced by the pro-Sternberg bias, while in the rate-changed condition the ratings are unaffected by the bias manipulation.

Single-task Tracking Dissociation. Dissociations between performance and subjective workload assessments as a result of a pay manipulation is one prediction of the multiple resource model (e.g., Wickens & Yeh, 1983). The logic behind this prediction is as follows: One, the availability of a bonus will increase the subject's motivation to perform well. Two, this increased motivation will lead to an increased mobilization of resources in general and increased allocation of resources to the specific

Table 5

Consistency x Type of Measure Interaction Means from

Z-Score Analysis

Analysis and Measure Type	Consistent Mapping Trials Mean	Varied Mapping Trials Mean
Perceptual-Load Analysis		
Task Difficulty	-0.03	0.28
Performance Rating	0.17	-0.27
RMS Error	0.03	0.01
Rate-Change Analysis		
Task Difficulty	-0.43	0.02
Overall Workload	-0.29	-0.02
Performance Rating	0.24	-0.09
RMS Error	-0.05	-0.04

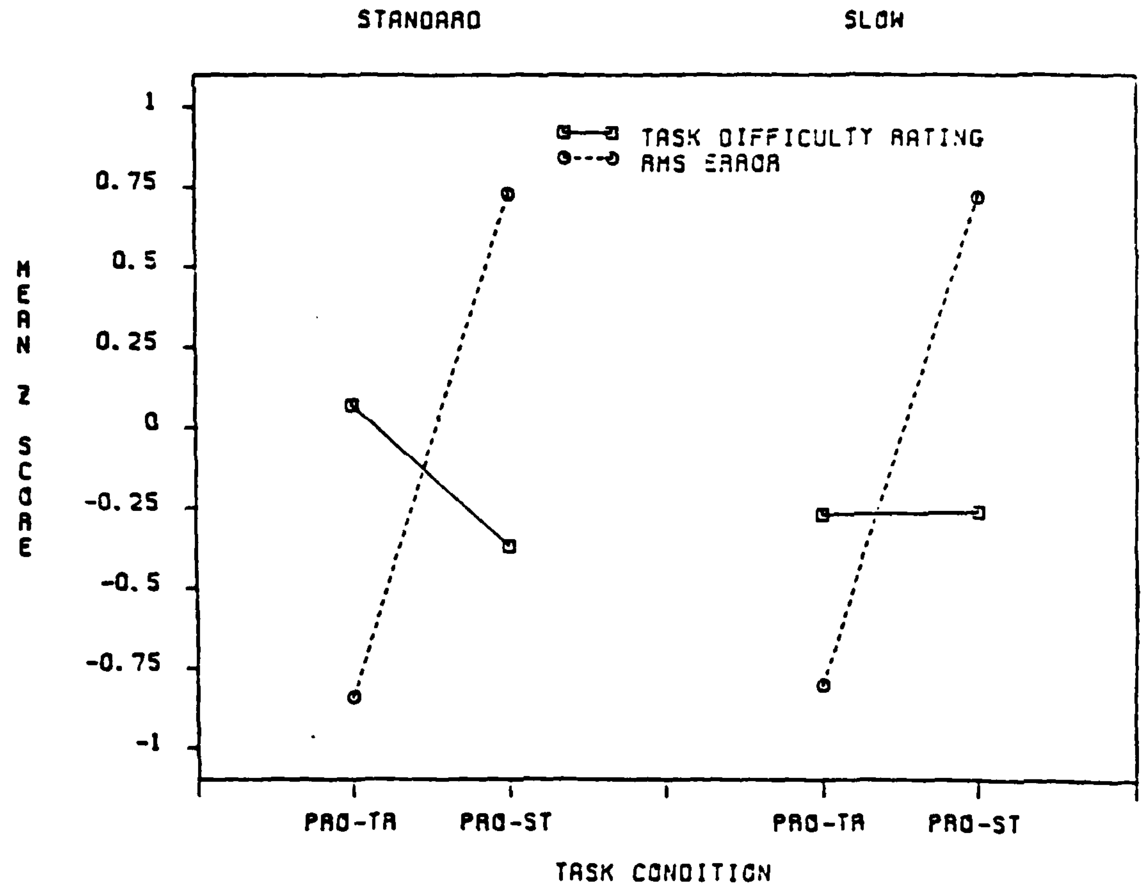

Figure 4. Rate-change x Pay x Type of Measure interaction in the *z*-score dissociation analysis of Task Difficulty rating and RMS error.

relevant task. Three, this increase in allocated resources will improve the performance in any resource-limited task, while being subjectively experienced as increased effort or workload. A very pure test of this hypothesis is provided by the single-task tracking data in which the only independent variable manipulation was the availability of the bonus.

The untransformed data from the overall RMS error and the eight rating scales were analysed via nine one-tailed *t*-tests. Three of the *t*-tests were significant at the $p < .05$ level: Mean overall RMS error was reduced from .099 to .094 (t (39) = 23.6, $p < .0001$), mean Incentive rating increased from 10.2 to 11.7 (t (39) = 3.2, $p < .003$), and mean Task Difficulty rating increased from 9.0 to 9.5 (t (39) = -1.7, $p < .05$) The increase in the Incentive rating confirms the effectiveness of the pay variable. The RMS error and the Task Difficulty ratings effects, displayed in Figure 5, are in complete accordance with the predictions of Wickens and Yeh (1983).

Discussion

The experimental results have important implications for the human factors practitioners involved in workload assessment. A number of dissociations were induced by the experimental manipulations. The dissociations illuminate the differences between the cognitive processing that generates subjective ratings and the cognitive processing that generates performance. This helps to outline limitations of the subjective assessment technique. Three experimental manipulations were effective in producing dissociations: consistency, rate-changing, and pay. Each of these will be discussed in turn.

Number of Tasks x Consistency Dissociation

The most dramatic dissociation is probably the failure of subjective assessments using the global scales to accurately reflect the dual-task advantage associated with the consistently mapped Sternberg configuration (refer to Figure 2). The differences in subjective assessments across the consistent and varied mapped Sternberg configurations on single-task trials agree well with the performance changes, but in the dual-task trials there is a marked performance advantage to the consistently mapped Sternberg configuration that is not apparent from the subjective workload assessments.

Apparently, the presence of the tracking task drove the subjective workload assessments to such a degree that the difference between the Sternberg configurations was diluted. This does not necessarily imply that the assessments are unreliable indicators of the subjectively experienced workload. Experientially, the presence or absence of the tracking task could be such a major contributor to the experience of workload that the consistent versus varied mapped distinction is trivial, even though the distinction between Sternberg configurations was distinct in the single-task trials. However, the presence of the

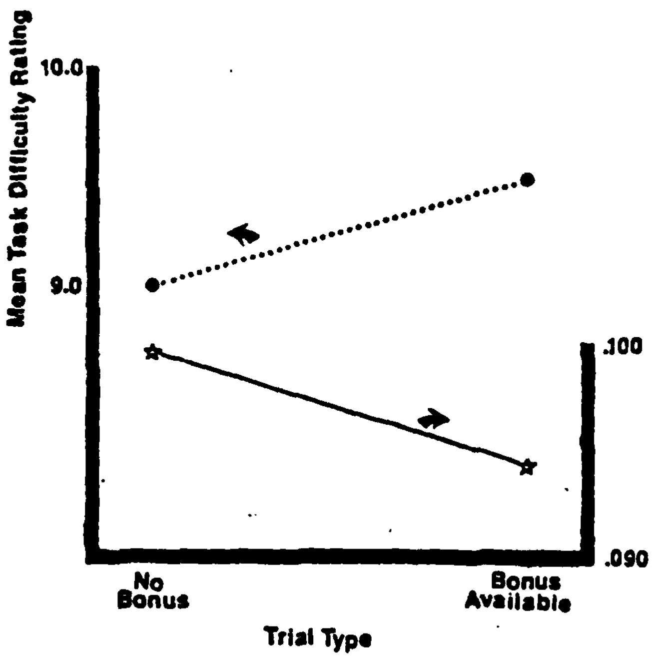

Figure 5. Pay dissociation from the single-task tracking trials.

tracking task did not eliminate the performance advantage enjoyed by the consistently mapped Sternberg configuration. This result can be viewed as another case of the dominating effect of the number of tasks to be performed has on subjective workload ratings. Several previous researchers have obtained similar results (e.g., Wickens & Derrick, 1981; Wickens & Yeh, 1982).

The implications of this finding to the applied practitioners of workload assessment are obvious. First, a multi-task environment may reduce the utility of subjective workload assessments to detect the advantages or disadvantages associated with reconfiguring one of the tasks. An applied situation similar to the one investigated in this study would be workload assessments collected in an aircraft cockpit. The basic control of an aircraft in flight or a flight simulator might be such a major contributor to the experiences and/or ratings of workload (especially in novice pilots) that different side-task configurations with very real differences in performance might be rated approximately the same. This implies that subjective assessments of such competing side-task configurations should, if possible, be gathered in the single-task environment as well as the multi-task and that small differences in the multi-task situation may need to be weighted more than similar differences in single-task configurations.

This evaluation might at first glance seem inconsistent with the conclusions of Wickens and Derrick (1981) and Wickens and Yeh (1982). These studies indicated that subjective workload assessments were relatively insensitive to increasing single-task difficulty. However, the ratings were insensitive relative to their ability to reflect the increase in demands evoked by adding a second task. The present findings indicate that when the issue of interest is a change in a single-task's difficulty, the efficiency of subjective ratings to detect that effect is degraded when combined with simultaneous performance of another task.

The results of this study indicate that this is the case at least when the change in single-task difficulty is induced by different levels of automaticity. It is worth noting that the dissociation was achieved with what could be considered a relatively mild level of automaticity. Automaticity develops with extended practice at detecting consistently mapped targets. In the entire experiment there were less than 2,000 opportunities to detect these targets. Much of the research in automaticity is based on many more trials (e.g., Schneider & Fisk, 1982a, 1982b). More extended training would be expected to increase the level of automaticity and quite probably the probability or size of a dissociation as well.

<u>Rate-change Dissociation</u>

A second important dissociation concerns the impact of the rate-change manipulation. Subjects rated the slower rate-changed trials easier, but their reaction times were slowed. One possible explanation would be differences in arousal between the two rate

of presentation conditions. The slower trials could conceivably have encouraged subjects to adopt a more relaxed attitude. This lower level of arousal could slow reaction times and ease the subjective experience of stress or workload. However, this explanation would suggest a Pay x Rate-change x Type of Measure interaction in the z-score dissociation analysis since a higher arousal level should be maintained by the bonus availability even on the slow rate-changed trials. There was no evidence of such an interaction present in the data.

Another, more plausible, explanation involves the development of subjects' expectancies. The subjects were trained most extensively in the standard Sternberg configuration in which the stimulus presentation rate was twice as fast as in the rate-changed condition. Consequently, the subjects developed a timing strategy, or rhythm, that was inappropriate to the rate-changed condition. As would be expected, this disruption of subject's expectancies increased the mean reaction time to the Sternberg stimuli. Inconsistent with this performance result is the finding that the subjective assessments of workload were reduced by the rate-change manipulation. Although the workload ratings findings are inconsistent with the performance effects, they are not surprising. The change in rate of the stimuli presentations is phenomenally very salient, and the reduction in speed of incoming stimuli is usually associated with lower workload. Relative to this mechanism, the disruption of temporal expectencies, as done in this experiment, could produce more subtle experiential effects.

This result is reminiscent of the inferential contamination mechanism suggested by Nisbett and Wilson's (1977) theory. Basically, this would suggest that the subjects are very aware of the rate of incoming stimuli and that lower rates are normally associated with less work. This logical analysis of the situation could override detection of the performance reducing effects of the disrupted rhythm.

Dealing with this mechanism of dissociation in an applied setting could be difficult since cataloging all the possible "logical" analyses which could bias subjects' ratings would be a prohibitively difficult undertaking. In some cases use of a between-subject design could help.

Pay Dissociation

A third dissociation occurred between the subjects' ratings of performance and their actual reaction time performance when the pay variable was manipulated. When there was a bonus available which was contingent on their Sternberg performance, subjects rated their performance lower but also had faster reaction times. This indicates that some sort of criterion shift occurs; when the bonus is available, subjects tend to become more critical of their performance in the bonus-available condition.

However, the data in Table 4 indicate that for the Sternberg task, this effect seems to be isolated to the dual-task condition where the pay manipulation changes the bias from one task to another, but a bonus was always available. Subjects consistently rated both the workload and their performance higher in the pro-tracking condition. This, of course, caused interactions when compared to the Sternberg task reaction time dependent variable.

Nevertheless, the motivation dissociation predicted by Wickens and Yeh (1983) was found in the single-task tracking data. The subjects rated their motivation during bonus-available trials higher on the incentive scale and this was associated with both improved performance and higher ratings of task difficulty. This finding supports the theoretical conjecture that higher levels of motivation increase the allocated resources which will improve performance on any resource-limited task, but will be perceived by subjects as increasing difficulty. To an applied worker this result indicates the need to maintain equivalent levels of motivation over groups of subjects and task conditions.

The fact that there were no reliable interactions involving the perceptual-loading manipulation is interesting. Given that perceptual loading was effective at changing performance this lack of dissociation might be related to the fact that manipulations that affect the "early" stages of processing are better suited for subjective assessments. This would be consistent with extrapolations from the Ericsson and Simon (1980) view that verbal reports are based primarily on activity in working memory.

Overall, the present study supports the previous findings of dissociations contaminating the interpretation of subjective workload assessments (e.g., Wickens & Derrick, 1981; Wickens & Yeh, 1982). There is also a suggestion that a number of underlying mechanisms may be at work. In the dissociation involving the consistency manipulation the results suggest that in a multi-task environment the changes in a single-task's difficulty might be too subtle to be reflected in the ratings. The dissociation resulting from the rate-change manipulation, on the other hand, suggests that subject's ratings can, on occasion, be contaminated by a "logical" analysis of task demands. Taken as a whole, the previous and present results suggest that basing evaluations of systems on subjective ratings alone could be risky. On the other hand, one should not focus so strongly on the dissociations as to lose sight of the fact that often the ratings were in good agreement with the performance effects.

Global versus Specific Scales

The choice of scales is crucial in this type of subjective workload asessment. An important consideration is to identify what scales can most efficiently provide accurate data. The term "accurate data" in this case refers to a scale which responds most like the objective performance data. While the factors contributing to the subjective experience of workload are intrinsically interesting, the important thing for applied

practitioners is to realize how these workload assessments relate to objective performance. One question raised by this issue is whether the scales used should be global or specific. Global scales usually attempt to answer the question that is most important to the practitioner (i.e., "which task is the most difficult or has the highest workload?"). But, clearly there are a multitude of environmental and organismic components to workload, and it seems logical to ask the subjects to distinguish between them. This led to the development of a multitude of specific scales. Quite a few contemporary approaches depend on the assumption that multiple scales are useful in isolating different components of the workload exprience. Sheridan and Simpson (1979) suggest that subjective workload is composed of three basic components: time pressure, complexity, and stress. Hauser, Childress, and Hart (1982) filtered through 15 scales looking for the best set. Eggemeier et. al (1982) used SWAT, a set of subjective ratings similar to Sheridan and Simpson's, in their work.

The results of the regressions predicting global scale ratings from combinations of specific scales were very encouraging. The combination of the Mental/Sensory Effort, Response Load, and Stress Level scales usually explained over half of the variance in the Overall Workload and Task Difficulty scales. This could be indicative of a relatively uniform concept of workload which is based on phenomenally salient components. Two of these potential components, Mental/Sensory Effort and Response Load, are consistent with what multiple resource theory (Wickens, 1980) would predict to be crucial. These two scales may measure resource competition within the two stages of processing postulated by the multiple resource theory.

However, the results do favor a certain amount of caution before using combinations of specific scales to assess workload. After all, the single most sensitive scale was Task Difficulty, a global scale. Also, some of the effects of the specific scale ratings were quite misleading. Most notable was the abysmal performance of the Time Pressure scale. Time pressure responded to only three of the independent manipulations and two of these effects were dissociations from the performance data. The weakness of this scale is particularly worrisome given the prominant role similar scales have in some subjective workload assessment methods (e.g., Sheridan & Simpson, 1979). The Response Load scale also displayed some questionable tendencies when it reacted to the consistency and perceptual-load manipulations; neither of which was expected to influence respose load. In general, the specific scales, especially Time Pressure, seemed particularly susceptible to inferential contamination and/or other biases.

There are at least two potential explanations for this: (1) subjects are unable to accurately distinguish the levels of such specific scales (i.e., the scales may represent a non-phenomenal component of the workload experience), or (2) the method used to collect these scale values is improper. Perhaps subjects could

make better use of these scales if there weren't so many to deal with at the end of every trial. The Ericsson and Simon (1980) model of verbal reports suggests that memory limitations are potentially very damaging to retrospective report accuracy. Another way to reduce memory load would be to use concurrent reports. Rehmann, Stein, and Rosenberg (1983) used a multiple button device to collect ratings during performance. They concluded such a procedure increased rating sensitivity. This would be consistent with Ericsson and Simon's (1980) evaluation of concurrent versus retrospective reports. Rehmann et. al used a global overall workload scale, but the same technique might be helpful in specfic scale applications.

In any case, the use of multiple specific scales with the conventional rating collection techniques appears to be a less efficient procedure than the use of global scales. The Task Difficulty scale was particularly promising.

Conclusions

In the final analysis the present study is seen to support the following conclusions of interest to the applied practitioner:

(1) The multi-task environment is capable of obscuring the differences between levels of difficulty of a single-task component, even when the differences are readily detectable in the single-task environment.

(2) Subjective Workload ratings do not always accurately reflect the performance advantage of automaticity, especially in the multi-task environment.

(3) Objective task evaluations may contaminate the ratings of workload. In situations where a logical analysis of the different task conditions could lead subjects to expect effects contrary to those that actually occur their ratings may reflect their expectations. The rate-change manipulation may be one example of this type of mechanism at work. Making use of between-subject designs might reduce some of the potential for this.

(4) Higher levels of motivation induce higher levels of performance, but also raise assessments of perceived difficulty. This indicates a need for maintaining equivalent levels of motivation over groups of subjects and differing task conditions.

(5) Unless the need for a specific scale can be specified a priori, the subjective analysis of workload is best served by global scales. This statement is justified by the confusing behavior of the specific scales, particularly Time Pressure.

The implications of this study to an engineering psychology researcher are less obvious but perhaps even more important.

First of all, the results support a processing characteristic approach to studying workload dissociations. Tying dissociations to processing phenomena (such as automaticity) offers greater generality than a simple cateloging of task effects.

Secondly, the results suggest that there is great need for research on the methods for collecting subjective workload ratings. If, as is generally believed, workload is multidimensional then it seems likely that some role for specific scales does exist. However, the present results indicate that the simple conventional technique of collecting ratings is not harnessing this potential.

Clearly, this is a field of research which is likely to remain active for many years.

References

Bird, K. L. Subjective rating scales as a workload assessment technique. Proceedings of the 17th Annual Conference on Manual Control, 1981.

Borg, G. Subjective aspects of physical and mental load. Ergonomics, 1978, 21, 215-220.

Cooper, G. E., and Harper, R. P. The use of pilot rating in the evaluation of aircraft handling qualities. Moffett Field, CA: NASA, Ames Research Center, NASA Technical Note No. TN D-5153, April 1969.

Derrick W. L. The relationship between processing resource and subjective dimensions of operator workload. Proceedings of the Human Factors Society - 25th Annual Meeting, 1981.

Eggemeier, F. T., Crabtree, M. S., Zingg, J. J., Reid, G. B., and Shingledecker, C. A. A subjective workload assessment in a memory update task. Proceedings of the Human Factors Society - 26th Annual Meeting, 1982.

Ericsson, K. A., and Simon, H. A. Verbal reports as data. Psychological Review, 1980, 87, 215-251.

Hauser, J. R., Childress, M. E., and Hart, S. G. Rating consistency and component saliance in subjective workload estimation. Paper presented at the 18th Annual Conference on Manual Control, Dayton, June, 1982.

Higgins, T. H. A systems engineering evaluation method for piloted aircraft and other man-operated vehicles and machines with hypothetical example of a system's evaluation. Washington, D.C.: Federal Aviation Administration Report No. FAA-R9-78-78, September 1979.

Johanssen, G., Moray, N., Pew, R., Rasmussen, J., and Wickens, C. Final report of the experimental group. In N. Moray (Ed.), Mental workload: Its theory and measurement. New York: Plenum Press, 1979.

Maverick production opposed by GAO. Aviation Week and Space Technology, July 19, 1983, pp. 43.

Myers, J. L. Fundamentals of experimental design. Boston: Allyn and Bacon, 1979.

Nisbett, R. E., and Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.

Rehmann, J. T., Stein, E. S., and Rosenberg, B. L. Subjective pilot workload assessment. Human Factors,

1983, 25, 297-307.

Schneider, W., and Fisk, A. D. Concurrent automatic and controlled visual search: Can processing occur without resource cost? Journal of Experimental Psychology: Learning, Memory, and Cognition, 1982, 8, 261-278. (a)

Schneider, W., and Fisk, A. D. Degree of consistent training: Improvements in search performance and automatic process development. Perception & Psychophysics, 1982, 31, 160-168. (b)

Sheridan, T. B., and Simpson, R. W. Toward the definition and measurement of the mental workload of transport pilots. Cambridge Massachusetts: M.I.T. Flight Transportation Laboratory Report, FTL Report R79-4, January 1979.

Shingledecker, C. A. Enhancing operator acceptance and noninterference in secondary task measures of workload. Proceedings of the Human Factors Society - 24th Annual Meeting, 1980.

Titchner, E. B. The schema of introspection. American Journal of Psychology, 1912, 23, 485-508.

Watson, J. B. Psychology as the behaviorist views it. Psychological Review, 1913, 20, 158-177.

Watson, J. B. Is thinking merely the action of the language mechanism? British Journal of Psychology, 1920, 11, 87-104.

Wewerinke, P. H., and Smit, J. A simulator study to investigate human operator workload. NATO AGARD Conference Proceedings, 1974, 146, 2-1 - 2-6.

Wickens, C. D. The structure of attentional resources. In R. Nickerson (Ed.), Attention and Performance VIII. Englewood Cliffs, NJ: Erlbaum, 1980.

Wickens, C. D., and Derrick, W. Workload measurement and multiple resources. IEEE 1981 Proceedings of the International Conference on Cybernetics and Society, 1981.

Wickens, C. D., and Yeh, Y. Y. The dissociation of subjective ratings and performance. IEEE 1982 Proceedings of the International Conference on Cybernetics and Society, 1982.

Wickens, C. D., and Yeh, Y. Y. The dissociation between subjective workload and performance: A multiple resource approach. Proceedings of the Human Factors Society -

27th Annual Meeting, 1983, in press.

Wierwille, W. W., and Williges, R. C. Survey and analysis of operator workload assessment techniques. Blacksburg, VA: Systemetrics Technical Report No. S-78-101, September 1978.

OFFICE OF NAVAL RESEARCH

Engineering Psychology Group
TECHNICAL REPORTS DISTRIBUTION LIST

CAPT Paul R. Chatelier
Office of the Deputy Under
Secretary of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D.C. 20301

Engineering Psychology Group
Office of Naval Research
Code 442 EP
Arlington, VA 22217 (2 cys)

Aviation & Aerospace Technology
Programs
Code 210
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

Physiology & Neuro Biology
Programs
Code 441NB
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

Manpower, Personnel &
Training Programs
Code 270
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

Information Sciences Division
Code 433
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

CDR Paul Girard
Code 250
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

Special Assistant for
Marine Corps Matters
Code 100M
Office of Naval Research
Arlington, VA 22217

Dr. J. Lester
ONR Detachment
495 Summer St.
Boston, MA 02210

CDR James Offutt
Officer-in-Charge
ONR Detachment
1030 E. Green St.
Pasadena, CA 91106

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D.C. 20375

Dr. Julie Hopson
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Dr. Robert G. Smith
Office of the Chief of Naval
Operations, OP987H
Personnel Logistics Plans
Washington, D.C. 20350

Human Factors Department
Code N-71
Naval Training Equipment Center
Orlando, FL 32813

CDR Norman E. Lane
Code N-7A
Naval Training Equipment Center
Orlando, FL 32813

Dr. Gary Poock
Operations Research Dept.
Naval Postgraduate School
Monterey, CA 93940

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D.C. 20380

Dr. L. Chmura
Naval Research Laboratory
Code 7592
Computer Sciences & Systems
Washington, D.C. 20375

Human Factors Technology
Administrator
Office of Naval Technology
Code MAT 0722
800 N. Quincy St.
Arlington, VA 22217

Commander
Naval Air Systems Command
Human Factors Programs
NAVAIR 334A
Washington, D.C. 20361

Mr. Norm Beck
Combat Control, Code 3512
Naval Underwater Systems Center
Newport, RI 02840

Commander
Naval Air Systems Command
Crew Station Design
NAVAIR 5313
Washington, D.C. 20361

Mr. Philip Andrews
Naval Sea Systems Command
NAVSEA 6L R2
Washington, D.C. 20362

Larry Olmstead
Naval Surface Weapons Ctr.
NSWC/DL
Code N-32
Dahlgren, VA 22448

CAPT Robert Biersner
Naval Medical R&D Command
Code 44
Naval Medical Ctr.
Bethesda, MD 20014

Dr. Arthur Bachrach
Behavioral Sciences Dept.
Naval Medical Research Institute
Bethesda, MD 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab.
Naval Submarine Base
Groton, CT 06340

LT Dennis McBride
Aerospace Psychology Department
Naval Aerospace Med. Res. Lab.
Pensacola, FL 32508

Dr. Wayne Zachary
Analytics, Inc.
2500 Maryland Rd.
Willow Grove, PA 19090

Dr. Robert Blanchard
Navy Personnel Research
& Development Ctr.
Command and Support Systems
San Diego, CA 92152

CDR J. Funaro
Human Factors Engr. Div.
Naval Air Development Ctr.
Warminster, PA 18974

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons
China Lake, CA 93555

Aircrew Systems Branch
Systems Engr. Test Directorate
U.S. Naval Air Test Center
Patuxent River, MD 20670

CDR C. Hutchins
Code 55
Naval Postgraduate School
Monterey, CA 93940

Office of the Chief of Naval
Operations (OP-115)
Washington, D.C. 20350

Dr. Edgar M. Johnson
Technical Director
U.S. Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Director, Organizations &
Systems Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Technical Director
U.S. Army Human Engr. Labs.
Aberdeen Proving Ground, MD 21005

U.S. Air Force Office of
Scientific Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D.C. 20332

Chief, Systems Engr. Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Dr. Earl Alluisi
Chief Scientists
AFHRL/CCN
Brooks AFB, TX 78235

Dr. Daniel Kahneman
University of British Columbia
Department of Psychology
Vancouver, BC V6T 1W5
Canada

Director, Human Factors Wing
Defense & Civil Institute of
Environmental Medicine
P.O. Box 2000
Downsview, Ontario M3M 3B9
Canada

Defense Technical Information Ctr.
Cameron Station, Bldg. 5
Alexandria, VA 22314 (12 cys)

Dr. M. Montemerlo
Human Factors & Simulation
Technology, RTE-6
NASA HQS
Washington, D.C. 20546

Dr. Amos Tversky
Dept. of Psychology
Stanford University
Stanford, CA 94305

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. T. B. Sheridan
Dept. of Mechanical Engr.
Massachusetts Institute of
Technology
Cambridge, MA 02139

Dr. Harry Snyder
Dept. of Industrial Engineering
Virginia Polytechnic Institute
& State University
Blacksburg, VA 24061

Dr. Robert T. Hennessy
NAS - National Research Council (COHF)
2101 Constitution Ave., N.W.
Washington, D.C. 20418

Dr. Amos Freedy
Perceptronics, Inc.
6271 Variel Ave.
Woodland Hills, CA 91364

Dr. Robert C. Williges
Dept. of Ind. Engr. & OR
Virginia Polytechnic Institute
& State College
130 Whittemore Hall
Blacksburg, VA 24061

Dr. Deborah Boehm-Davis
General Electric Co.
Information Systems Programs
1755 Jefferson Davis Highway
Arlington, VA 22202

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Boulder, CO 80309

Dr. James H. Howard, Jr.
Dept. of Psychology
Catholic University
Washington, D.C. 20064

Dr. William Howell
Dept. of Psychology
Rice University
Houston, TX 77001

Dr. Edward R. Jones
Chief, Human Factors Engr.
McDonnell-Douglas Astronautics Co.
St. Louis Division
Box 516
St. Louis, MO 63166

Dr. Babur M. Pulat
Dept. of Industrial Engineering
North Carolina A&T State Univ.
Greensboro, NC 27411

Dr. Lola Lopes
Information Sciences Division
Dept. of Psychology
University of Wisconsin
Madison, WI 53706

Dr. Stanley N. Roscoe
New Mexico State Univ.
Box 5095
Las Cruces, NM 88003

Dr. William B. Rouse
School of Industrial &
Systems Engr.
Georgia Institute of Technology
Atlanta, GA 30332

Dr. Richard Pew
Bolt Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02238

Dr. Hillel Einhorn
Graduate School of Business
University of Chicago
1101 E. 58th St.
Chicago, IL 60637

Dr. Douglas Towne
University of Southern California
Behavioral Technology Laboratory
3716 S. Hope St.
Los Angeles, CA 90007

Dr. Marshall Farr
Office of Naval Research
Code 442
800 N. Quincy St.
Arlington, VA 22217

Mr. Joseph G. Wohl
Alphatech Inc.
2 Burlington Executive Ctr.
111 Middlesex Turnpike
Burlington, MA 01803

END

FILMED

11-83

DTIC